

An ω -Power of a Finitary Language Which is a Borel Set of Infinite Rank

Olivier Finkel

Equipe de Logique Mathématique

U.F.R. de Mathématiques, Université Paris 7

2 Place Jussieu 75251 Paris cedex 05, France.

E Mail: finkel@logique.jussieu.fr

Abstract

ω -powers of finitary languages are ω -languages in the form V^ω , where V is a finitary language over a finite alphabet Σ . Since the set Σ^ω of infinite words over Σ can be equipped with the usual Cantor topology, the question of the topological complexity of ω -powers naturally arises and has been raised by Niwinski [Niw90], by Simonnet [Sim92], and by Staiger [Sta97b]. It has been proved in [Fin01] that for each integer $n \geq 1$, there exist some ω -powers of context free languages which are Π_n^0 -complete Borel sets, and in [Fin03] that there exists a context free language L such that L^ω is analytic but not Borel. But the question was still open whether there exists a finitary language V such that V^ω is a Borel set of infinite rank.

We answer this question in this paper, giving an example of a finitary language whose ω -power is Borel of infinite rank.

Keywords: Infinite words; ω -languages; ω -powers; Cantor topology; topological complexity; Borel sets; infinite rank.

1 Introduction

ω -powers are ω -languages in the form V^ω , where V is a finitary language. The operation $V \rightarrow V^\omega$ is a fundamental operation over finitary languages leading to ω -languages. This operation appears in the characterization of the class REG_ω of ω -regular languages (respectively, of the class CF_ω of

context free ω -languages) as the ω -Kleene closure of the family REG of regular finitary languages (respectively, of the family CF of context free finitary languages) [Sta97a].

The set Σ^ω of infinite words over a finite alphabet Σ is usually equipped with the Cantor topology which may be defined by a distance, see [Sta97a] [PP04]. One can then study the complexity of ω -languages, i.e. languages of infinite words, by considering their topological complexity, with regard to the Borel hierarchy (and beyond to the projective hierarchy) [Sta97a] [PP04].

The question of the topological complexity of ω -powers of finitary languages naturally arises. It has been posed by Niwinski [Niw90], by Simonnet [Sim92] and by Staiger [Sta97b]. The ω -power of a finitary language is always an analytic set because it is the continuous image of a compact set $\{0, 1, \dots, n\}^\omega$ for $n \geq 0$ or of the Baire space ω^ω , [Sim92] [Fin01]. It has been proved in [Fin01] that for each integer $n \geq 1$, there exist some ω -powers of context free languages which are Π_n^0 -complete Borel sets, and in [Fin03] that there exists a context free language L such that L^ω is analytic but not Borel.

But the question was still open whether there exists a finitary language V such that V^ω is a Borel set of infinite rank.

We answer this question in this paper, giving an example of a finitary language whose ω -power is Borel of infinite rank.

The paper is organized as follows. In Section 2 we recall definitions of Borel sets and previous results and we proved our main result in Section 3.

2 Recall on Borel sets and previous results

We assume the reader to be familiar with the theory of formal ω -languages [Tho90], [Sta97a]. We shall use usual notations of formal language theory.

When Σ is a finite alphabet, a *non-empty finite word* over Σ is any sequence $x = a_1 \dots a_k$, where $a_i \in \Sigma$ for $i = 1, \dots, k$, and k is an integer ≥ 1 . The *length* of x is k , denoted by $|x|$. The *empty word* has no letter and is denoted by λ ; its length is 0. For $x = a_1 \dots a_k$, we write $x(i) = a_i$ and $x[i] = x(1) \dots x(i)$ for $i \leq k$ and $x[0] = \lambda$. Σ^* is the *set of finite words* (including the empty word) over Σ .

The *first infinite ordinal* is ω . An ω -word over Σ is an ω -sequence $a_1 \dots a_n \dots$, where for all integers $i \geq 1$, $a_i \in \Sigma$. When σ is an ω -word over Σ , we write $\sigma = \sigma(1)\sigma(2) \dots \sigma(n) \dots$, where for all i , $\sigma(i) \in \Sigma$, and $\sigma[n] = \sigma(1)\sigma(2) \dots \sigma(n)$ for all $n \geq 1$ and $\sigma[0] = \lambda$.

The *prefix relation* is denoted \sqsubseteq : a finite word u is a *prefix* of a finite word v (respectively, an infinite word v), denoted $u \sqsubseteq v$, if and only if there exists a finite word w (respectively, an infinite word w), such that $v = u.w$. The *set of ω -words* over the alphabet Σ is denoted by Σ^ω . An *ω -language* over an alphabet Σ is a subset of Σ^ω .

For $V \subseteq \Sigma^*$, the *ω -power* of V is the ω -language:

$$V^\omega = \{\sigma = u_1 \dots u_n \dots \in \Sigma^\omega \mid \forall i \geq 1 \ u_i \in V - \{\lambda\}\}$$

We assume the reader to be familiar with basic notions of topology which may be found in [Mos80] [LT94] [Kec95] [Sta97a] [PP04]. For a finite alphabet X , we consider X^ω as a topological space with the Cantor topology. The *open sets* of X^ω are the sets of the form $W.X^\omega$, where $W \subseteq X^*$. A set $L \subseteq X^\omega$ is a *closed set* iff its complement $X^\omega - L$ is an open set. Define now the *Borel Hierarchy* of subsets of X^ω :

Definition 2.1 For a non-null countable ordinal α , the classes Σ_α^0 and Π_α^0 of the Borel Hierarchy on the topological space X^ω are defined as follows:

Σ_1^0 is the class of open subsets of X^ω .

Π_1^0 is the class of closed subsets of X^ω .

and for any countable ordinal $\alpha \geq 2$:

Σ_α^0 is the class of countable unions of subsets of X^ω in $\cup_{\gamma < \alpha} \Pi_\gamma^0$.

Π_α^0 is the class of countable intersections of subsets of X^ω in $\cup_{\gamma < \alpha} \Sigma_\gamma^0$.

For a countable ordinal α , a subset of X^ω is a Borel set of *rank* α iff it is in $\Sigma_\alpha^0 \cup \Pi_\alpha^0$ but not in $\cup_{\gamma < \alpha} (\Sigma_\gamma^0 \cup \Pi_\gamma^0)$.

There are also some subsets of X^ω which are not Borel. In particular the class of Borel subsets of X^ω is strictly included into the class Σ_1^1 of *analytic sets* which are obtained by projection of Borel sets, see for example [Sta97a] [LT94] [PP04] [Kec95] for more details.

We now define completeness with regard to reduction by continuous functions. For a countable ordinal $\alpha \geq 1$, a set $F \subseteq X^\omega$ is said to be a Σ_α^0 (respectively, Π_α^0 , Σ_1^1)-*complete set* iff for any set $E \subseteq Y^\omega$ (with Y a finite alphabet): $E \in \Sigma_\alpha^0$ (respectively, $E \in \Pi_\alpha^0$, $E \in \Sigma_1^1$) iff there exists a continuous function $f : Y^\omega \rightarrow X^\omega$ such that $E = f^{-1}(F)$. Σ_n^0 (respectively Π_n^0)-complete sets, with n an integer ≥ 1 , are thoroughly characterized in [Sta86].

In particular $\mathcal{R} = (0^*.1)^\omega$ is a well known example of Π_2^0 -complete subset of $\{0, 1\}^\omega$. It is the set of ω -words over $\{0, 1\}$ having infinitely many occurrences of the letter 1. Its complement $\{0, 1\}^\omega - (0^*.1)^\omega$ is a Σ_2^0 -complete subset of $\{0, 1\}^\omega$.

We shall recall the definition of the operation $A \rightarrow A^\approx$ over sets of infinite words we introduced in [Fin01] and which is a simple variant of Duparc's operation of exponentiation $A \rightarrow A^\sim$ [Dup01].

For a finite alphabet Σ we denote $\Sigma^{\leq \omega} = \Sigma^\omega \cup \Sigma^*$. Let now \leftarrow a letter not in Σ and $X = \Sigma \cup \{\leftarrow\}$. For $x \in X^{\leq \omega}$, x^\leftarrow denotes the string x , once every \leftarrow occurring in x has been “evaluated” to the back space operation (the one familiar to your computer!), proceeding from left to right inside x . In other words $x^\leftarrow = x$ from which every interval of the form “ $a \leftarrow$ ” ($a \in \Sigma$) is removed. We add the convention that $(u. \leftarrow)^\leftarrow$ is undefined if $|u^\leftarrow| = 0$, i.e. when the last letter \leftarrow can not be used as an eraser (because every letter of Σ in u has already been erased by some erasers \leftarrow placed in u). Remark that the resulting word x^\leftarrow may be finite or infinite.

For example if $u = (a \leftarrow)^n$, for $n \geq 1$, or $u = (a \leftarrow)^\omega$ then $(u)^\leftarrow = \lambda$, if $u = (ab \leftarrow)^\omega$ then $(u)^\leftarrow = a^\omega$, if $u = bb(\leftarrow a)^\omega$ then $(u)^\leftarrow = b$, if $u = \leftarrow (a \leftarrow)^\omega$ or $u = a \leftarrow \leftarrow a^\omega$ or $u = (a \leftarrow \leftarrow)^\omega$ then $(u)^\leftarrow$ is undefined.

Definition 2.2 For $A \subseteq \Sigma^\omega$, $A^\approx = \{x \in (\Sigma \cup \{\leftarrow\})^\omega \mid x^\leftarrow \in A\}$.

The following result follows easily from [Dup01] and was applied in [Fin01] to study the ω -powers of finitary context free languages.

Theorem 2.3 Let n be an integer ≥ 2 and $A \subseteq \Sigma^\omega$ be a Π_n^0 -complete set. Then A^\approx is a Π_{n+1}^0 -complete subset of $(\Sigma \cup \{\leftarrow\})^\omega$.

For each ω -language $A \subseteq \Sigma^\omega$, the ω -language A^\approx can be easily described from A by the use of the notion of substitution which we recall now.

A *substitution* is defined by a mapping $f : \Sigma \rightarrow P(\Gamma^*)$, where $\Sigma = \{a_1, \dots, a_n\}$ and Γ are two finite alphabets, $f : a_i \rightarrow L_i$ where for all integers $i \in [1; n]$, $f(a_i) = L_i$ is a finitary language over the alphabet Γ .

Now this mapping is extended in the usual manner to finite words: $f(a_{i_1} \dots a_{i_n}) = L_{i_1} \dots L_{i_n}$, and to finitary languages $L \subseteq \Sigma^*$: $f(L) = \cup_{x \in L} f(x)$. If for each integer $i \in [1; n]$ the language L_i does not contain the empty word, then the mapping f may be extended to ω -words:

$$f(x(1) \dots x(n) \dots) = \{u_1 \dots u_n \dots \mid \forall i \geq 1 \quad u_i \in f(x(i))\}$$

and to ω -languages $L \subseteq \Sigma^\omega$ by setting $f(L) = \cup_{x \in L} f(x)$.

Let $L_1 = \{w \in (\Sigma \cup \{\leftarrow\})^* \mid w^{\leftarrow} = \lambda\}$. L_1 is a context free (finitary) language generated by the context free grammar with the following production rules: $S \rightarrow aS \leftarrow S$ with $a \in \Sigma$; and $S \rightarrow \lambda$ (where λ is the empty word).

Then, for each ω -language $A \subseteq \Sigma^\omega$, the ω -language $A^\approx \subseteq (\Sigma \cup \{\leftarrow\})^\omega$ is obtained by substituting in A the language $L_1.a$ for each letter $a \in \Sigma$.

By definition the operation $A \rightarrow A^\approx$ conserves the ω -powers of finitary languages. Indeed if $A = V^\omega$ for some language $V \subseteq \Sigma^*$ then $A^\approx = g(V^\omega) = (g(V))^\omega$ where $g : \Sigma \rightarrow P((\Sigma \cup \{\leftarrow\})^*)$ is the substitution defined by $g(a) = L_1.a$ for every letter $a \in \Sigma$.

3 An ω -power which is Borel of infinite rank

We can now iterate k times this operation $A \rightarrow A^\approx$. More precisely, we define, for a set $A \subseteq \Sigma^\omega$:

$$\begin{aligned} A_k^{\approx,0} &= A, \\ A_k^{\approx,1} &= A^\approx, \\ A_k^{\approx,2} &= (A_k^{\approx,1})^\approx, \text{ and} \\ A_k^{\approx,k} &= (A_k^{\approx,(k-1)})^\approx, \end{aligned}$$

where we apply k times the operation $A \rightarrow A^\approx$ with different new letters $\leftarrow_k, \leftarrow_{k-1}, \dots, \leftarrow_3, \leftarrow_2, \leftarrow_1$, in such a way that we have successively:

$$\begin{aligned} A_k^{\approx,0} &= A \subseteq \Sigma^\omega, \\ A_k^{\approx,1} &\subseteq (\Sigma \cup \{\leftarrow_k\})^\omega, \\ A_k^{\approx,2} &\subseteq (\Sigma \cup \{\leftarrow_k, \leftarrow_{k-1}\})^\omega, \\ &\dots\dots\dots \\ A_k^{\approx,k} &\subseteq (\Sigma \cup \{\leftarrow_k, \leftarrow_{k-1}, \dots, \leftarrow_1\})^\omega. \end{aligned}$$

For a reason which will be clear later we have chosen to successively call the erasers $\leftarrow_k, \leftarrow_{k-1}, \dots, \leftarrow_2, \leftarrow_1$, in this precise order. We set now $A^{\approx,k} = A_k^{\approx,k}$ so it holds that

$$A^{\approx,k} \subseteq (\Sigma \cup \{\leftarrow_k, \leftarrow_{k-1}, \dots, \leftarrow_1\})^\omega$$

Notice that definitions of $A_k^{\approx,1}, A_k^{\approx,2}, \dots, A_k^{\approx,(k-1)}$ were just some intermediate steps for the definition of $A^{\approx,k}$ and will not be used later.

We can also describe the operation $A \rightarrow A^{\approx,k}$ in a similar manner as in the case of the operation $A \rightarrow A^\approx$, by the use of the notion of substitution.

Let $L_k \subseteq (\Sigma \cup \{\leftarrow_k, \leftarrow_{k-1}, \dots, \leftarrow_1\})^*$ be the language containing (finite) words u , such that all letters of u have been erased when the operations of erasing using the erasers $\leftarrow_1, \leftarrow_2, \dots, \leftarrow_{k-1}, \leftarrow_k$, are successively applied to u .

Notice that the operations of erasing have to be done in a good order: the first operation of erasing uses the eraser \leftarrow_1 , then the second one uses the eraser \leftarrow_2 , and so on \dots

So an eraser \leftarrow_j may only erase a letter of Σ or an other “eraser” \leftarrow_i for some integer $i > j$.

It is easy to see that L_k is a context free language. In fact L_k belongs to the subclass of iterated counter languages which is the closure under substitution of the class of one counter languages, see [ABB96] [Fin01] for more details.

Let now h_k be the substitution: $\Sigma \rightarrow P((\Sigma \cup \{\leftarrow_k, \leftarrow_{k-1}, \dots, \leftarrow_1\})^*)$ defined by $h_k(a) = L_k.a$ for every letter $a \in \Sigma$.

Then it holds that, for $A \subseteq \Sigma^\omega$, $A^{\approx.k} = h_k(A)$, i.e. $A^{\approx.k}$ is obtained by substituting in A the language $L_k.a$ for each letter $a \in \Sigma$.

The ω -language $\mathcal{R} = (0^*.1)^\omega = \mathcal{V}^\omega$, where $\mathcal{V} = (0^*.1)$, is $\mathbf{\Pi}_2^0$ -complete. Then, by Theorem 2.3, for each integer $p \geq 1$, $h_p(\mathcal{V}^\omega) = (h_p(\mathcal{V}))^\omega$ is a $\mathbf{\Pi}_{p+2}^0$ -complete set.

We can see that the languages L_k , for $k \geq 1$, form a sequence which is strictly increasing for the inclusion relation:

$$L_1 \subset L_2 \subset L_3 \subset \dots \subset L_i \subset L_{i+1} \dots$$

In order to construct some ω -power which is Borel of infinite rank, we could try to substitute the language $\cup_{k \geq 1} L_k.a$ to each letter $a \in \Sigma$. But the language $\cup_{k \geq 1} L_k.a$ is defined over the infinite alphabet $\Sigma \cup \{\leftarrow_1, \leftarrow_2, \leftarrow_3, \dots\}$, so we shall first code every eraser \leftarrow_j by a finite word over a fixed finite alphabet. The eraser \leftarrow_j will be coded by the finite word $\alpha.\beta^j.\alpha$ over the alphabet $\{\alpha, \beta\}$, where α and β are two new letters.

The morphism $\varphi_p : (\Sigma \cup \{\leftarrow_1, \dots, \leftarrow_p\})^* \rightarrow (\Sigma \cup \{\alpha, \beta\})^*$ defined by $\varphi_p(c) = c$ for each $c \in \Sigma$ and $\varphi_p(\leftarrow_j) = \alpha.\beta^j.\alpha$ for each integer $j \in [1, p]$, can be naturally extended to a continuous function $\psi_p : (\Sigma \cup \{\leftarrow_1, \dots, \leftarrow_p\})^\omega \rightarrow (\Sigma \cup \{\alpha, \beta\})^\omega$.

Let now

$$\mathcal{L} = \bigcup_{n \geq 1} \varphi_n(L_n)$$

and $h : \Sigma \rightarrow P((\Sigma \cup \{\alpha, \beta\})^*)$ be the substitution defined by

$$h(a) = \mathcal{L}.a$$

for each $a \in \Sigma$.

We can now state our main result:

Theorem 3.1 *Let $\mathcal{V} = (0^*.1)$. Then the ω -power $(h(\mathcal{V}))^\omega \subseteq \{0, 1, \alpha, \beta\}^\omega$ is a Borel set of infinite rank.*

To prove this theorem, we shall proceed by successive lemmas.

Lemma 3.2 *For all integers $p \geq 1$, the ω -language $\psi_p(\mathcal{R}^{\approx.p})$ is a $\Pi_{\mathbf{p}+2}^0$ -complete subset of $(\Sigma \cup \{\alpha, \beta\})^\omega$.*

Proof. First we prove that $\psi_p(\mathcal{R}^{\approx.p})$ is in the class $\Pi_{\mathbf{p}+2}^0$.

For $\Sigma = \{0, 1\}$, $\psi_p((\Sigma \cup \{\leftarrow_1, \dots, \leftarrow_p\})^\omega)$ is the continuous image by ψ_p of the compact set $(\Sigma \cup \{\leftarrow_1, \dots, \leftarrow_p\})^\omega$, hence it is also a compact set.

The function ψ_p is injective and continuous thus it induces an homeomorphism ψ'_p between the two compact sets $(\Sigma \cup \{\leftarrow_1, \dots, \leftarrow_p\})^\omega$ and $\psi_p((\Sigma \cup \{\leftarrow_1, \dots, \leftarrow_p\})^\omega)$.

We have already seen that, for each integer $p \geq 1$, the ω -language $\mathcal{R}^{\approx.p}$ is a $\Pi_{\mathbf{p}+2}^0$ -complete subset of $(\Sigma \cup \{\leftarrow_1, \dots, \leftarrow_p\})^\omega$. Then $\psi'_p(\mathcal{R}^{\approx.p})$ is a $\Pi_{\mathbf{p}+2}^0$ -subset of $\psi_p((\Sigma \cup \{\leftarrow_1, \dots, \leftarrow_p\})^\omega)$, because the function ψ'_p is an homeomorphism.

But one can prove, by induction over the integer $j \geq 1$, that each Π_j^0 subset K of $\psi_p((\Sigma \cup \{\leftarrow_1, \dots, \leftarrow_p\})^\omega)$ is also a Π_j^0 subset of $(\Sigma \cup \{\alpha, \beta\})^\omega$. Thus $\psi'_p(\mathcal{R}^{\approx.p}) = \psi_p(\mathcal{R}^{\approx.p})$ is a $\Pi_{\mathbf{p}+2}^0$ -subset of $(\Sigma \cup \{\alpha, \beta\})^\omega$.

Remark now that the set $\mathcal{R}^{\approx.p}$ being $\Pi_{\mathbf{p}+2}^0$ -complete, every $\Pi_{\mathbf{p}+2}^0$ -subset of X^ω , for X a finite alphabet, is the inverse image of $\mathcal{R}^{\approx.p}$ by a continuous function. But it holds that $\mathcal{R}^{\approx.p} = \psi_p^{-1}(\psi_p(\mathcal{R}^{\approx.p}))$, where ψ_p is a continuous function. Thus every $\Pi_{\mathbf{p}+2}^0$ -subset of X^ω , for X a finite alphabet, is the inverse image of $\psi_p(\mathcal{R}^{\approx.p})$ by a continuous function. Therefore $\psi_p(\mathcal{R}^{\approx.p})$ is also a $\Pi_{\mathbf{p}+2}^0$ -complete subset of $(\Sigma \cup \{\alpha, \beta\})^\omega$. \square

Lemma 3.3 *The set $(h(\mathcal{V}))^\omega$ is not a Borel set of finite rank.*

Proof. Consider, for each integer $p \geq 1$, the regular ω -language

$$R_p = \psi_p(\{0, 1, \leftarrow_1, \leftarrow_2, \dots, \leftarrow_p\}^\omega) = \{0, 1, \alpha.\beta.\alpha, \alpha.\beta^2.\alpha, \dots, \alpha.\beta^p.\alpha\}^\omega$$

We have seen that R_p is compact hence it is a closed set. And by construction it holds that $(h(\mathcal{V}))^\omega \cap R_p = \psi_p(h_p(\mathcal{V}^\omega)) = \psi_p(\mathcal{R}^{\approx p})$ where $\mathcal{R} = \mathcal{V}^\omega$, so by Lemma 3.2 this set is a Π_{p+2}^0 -complete subset of $\{0, 1, \alpha, \beta\}^\omega$.

If $(h(\mathcal{V}))^\omega$ was a Borel set of finite rank it would be in the class Π_J^0 for some integer $J \geq 1$. But then $(h(\mathcal{V}))^\omega \cap R_p$ would be the intersection of a Π_J^0 -set and of a closed, i.e. Π_1^0 -set. Thus, for each integer $p \geq 1$, the set $(h(\mathcal{V}))^\omega \cap R_p$ would be a Π_J^0 -set. This would lead to a contradiction because, for $J = p$, a Π_J^0 -set cannot be a Π_{p+2}^0 -complete set. \square

Lemma 3.4 *Every ω -word $x \in (h(\mathcal{V}))^\omega$ has a unique decomposition of the form $x = u_1.u_2 \dots u_n \dots$ where for all $i \geq 1$ $u_i \in h(\mathcal{V})$.*

Proof. Towards a contradiction assume on the contrary that some ω -word $x \in (h(\mathcal{V}))^\omega = (h(0^*.1))^\omega$ has (at least) two distinct decompositions in words of $h(\mathcal{V})$. So there are some words $u_j, u'_j \in h(\mathcal{V})$, for $j \geq 1$, such that

$$x = u_1.u_2 \dots u_n \dots = u'_1.u'_2 \dots u'_n \dots$$

and an integer $J \geq 1$ such that $u_j = u'_j$ for $j < J$ and $u_J \sqsubset u'_J$, i.e. u_J is a strict prefix of u'_J . Then for each integer $j \geq 1$, there are integers $n_j, n'_j \geq 0$ such that $u_j \in h(0^{n_j}.1)$ and $u'_j \in h(0^{n'_j}.1)$. Thus there are some finite words $v_i^j \in \mathcal{L}$, where i is an integer in $[1, n_j + 1]$, and $w_i^j \in \mathcal{L}$, where i is an integer in $[1, n'_j + 1]$, such that

$$u_j = v_1^j.0.v_2^j.0 \dots v_{n_j}^j.0.v_{n_j+1}^j.1 \quad \text{and} \quad u'_j = w_1^j.0.w_2^j.0 \dots w_{n'_j}^j.0.w_{n'_j+1}^j.1$$

We consider now x given by its first decomposition $x = u_1.u_2 \dots u_n \dots$.

Let now $x^{(1)}$ be the ω -word obtained from x by using the (code of the) eraser \leftarrow_1 as an eraser which may erase letters 0, 1, and (codes of the) erasers \leftarrow_p for $p > 1$. Remark that by construction these operations of erasing occur inside the words v_i^j for $j \geq 1$ and $i \in [1, n_j + 1]$.

Next let $x^{(2)}$ be the ω -word obtained from $x^{(1)}$ by using the (code of the) eraser \leftarrow_2 as an eraser which may erase letters 0, 1, and (codes of the) erasers \leftarrow_p for $p > 2$. Again these erasing operations occur inside the words v_i^j .

We can now iterate this process. Assume that, after having successively used the erasers $\leftarrow_1, \leftarrow_2, \dots, \leftarrow_n$, for some integer $n \geq 1$, we have got the ω -word

$x^{(n)}$ from the ω -word x . We can now define $x^{(n+1)}$ as the ω -word obtained from $x^{(n)}$ by using the (code of the) eraser \leftarrow_{n+1} as an eraser which may erase letters 0, 1, and (codes of the) erasers \leftarrow_p for $p > n + 1$.

We shall denote $K_{(i,j)} = \min\{k \geq 1 \mid v_i^j \in \varphi_k(L_k)\}$. Then $v_i^j \in \varphi_{K_{(i,j)}}(L_{K_{(i,j)}})$ for all integers $j \geq 1$ and $i \in [1, n_j + 1]$. Thus after $K_j = \max\{K_{(i,j)} \mid i \in [1, n_j + 1]\}$ steps all words v_i^j , for $i \in [1, n_j + 1]$, have been completely erased and, from the finite word u_j , it remains only the finite word $0^{n_j}.1$.

After $K^J = \max\{K_j \mid j \in [1, J]\}$ steps, from the word $u_1.u_2 \dots u_J$, it remains the word $0^{n_1}.1.0^{n_2}.1.0^{n_3}.1 \dots 0^{n_J}.1$ which is a strict prefix of $x^{(K^J)}$. In particular the J -th letter 1 of $x^{(K^J)}$ is the last letter of u_J , which has not been erased, and it will not be erased by next erasing operations.

Notice that the successive erasing operations are in fact applied to x independently of the decomposition of x in words of $h(\mathcal{V})$. So consider now the above erasing operations applied to x given by its second decomposition.

Let $K'_{(i,j)} = \min\{k \geq 1 \mid w_i^j \in \varphi_k(L_k)\}$, and $K'^J = \max\{K'_{(i,j)} \mid j \in [1, J] \text{ and } i \in [1, n'_j + 1]\}$. $K^j = K'^j$ holds for $1 \leq j < J$ and $K^J \leq K'^J$.

We see that, after K'^J steps, the word $0^{n_1}.1.0^{n_2}.1.0^{n_3}.1 \dots 0^{n_{J-1}}.1.0^{n'_J}.1$ is a strict prefix of $x^{(K'^J)}$. The J -th letter 1 of $x^{(K'^J)}$ is the last letter of u'_J (which has not been erased). But we have seen above that it is also the last letter of u_J (which has not been erased).

Thus we would have $u_J = u'_J$ and this leads to a contradiction. \square

Remark that in terms of code theory Lemma 3.4 states that the language $h(\mathcal{V})$ is an ω -code.

Lemma 3.5 *The set $(h(\mathcal{V}))^\omega$ is a Borel set.*

Proof. Assume on the contrary that $(h(\mathcal{V}))^\omega$ is an analytic but non Borel set. Recall that lemma 4.1 of [FS03] states that if X and Y are finite alphabets having at least two letters and B is a Borel subset of $X^\omega \times Y^\omega$ such that $PROJ_{X^\omega}(B) = \{\sigma \in X^\omega \mid \exists \nu (\sigma, \nu) \in B\}$ is not Borel, then there are 2^{\aleph_0} ω -words $\sigma \in X^\omega$ such that $B_\sigma = \{\nu \in Y^\omega \mid (\sigma, \nu) \in B\}$ has cardinality 2^{\aleph_0} (where 2^{\aleph_0} is the cardinal of the continuum).

We can now reason as in the proof of Fact 4.5 in [FS03].

Let θ be a recursive enumeration of the set $h(\mathcal{V})$. The function $\theta : \mathbb{N} \rightarrow h(\mathcal{V})$ is a bijection and we denote $u_i = \theta(i)$.

Let now \mathcal{D} be the set of pairs $(\sigma, \nu) \in \{0, 1\}^\omega \times \{0, 1, \alpha, \beta\}^\omega$ such that:

1. $\sigma \in (0^*.1)^\omega$, so σ may be written in the form

$$\sigma = 0^{n_1}.1.0^{n_2}.1.0^{n_3}.1 \dots 0^{n_p}.1.0^{n_{p+1}}.1 \dots$$

where $\forall i \geq 1 \quad n_i \geq 0$, and

- 2.

$$\nu = u_{n_1}.u_{n_2}.u_{n_3} \dots u_{n_p}.u_{n_{p+1}} \dots$$

\mathcal{D} is a Borel subset of $\{0,1\}^\omega \times \{0,1,\alpha,\beta\}^\omega$ because it is accepted by a deterministic Turing machine with a Büchi acceptance condition [Sta97a].

But $PROJ_{\{0,1,\alpha,\beta\}^\omega}(\mathcal{D}) = (h(\mathcal{V}))^\omega$ would be a non Borel set thus there would be 2^{\aleph_0} ω -words ν in $(h(\mathcal{V}))^\omega$ such that \mathcal{D}_ν has cardinality 2^{\aleph_0} . This means that there would exist 2^{\aleph_0} ω -words $\nu \in (h(\mathcal{V}))^\omega$ having 2^{\aleph_0} decompositions in words in $h(\mathcal{V})$.

This would lead to a (strong) contradiction with Lemma 3.4. \square

Theorem 3.1 follows now directly from Lemmas 3.3 and 3.5.

4 Concluding Remarks

We already knew that there are ω -powers of every finite Borel rank [Fin01]. We have proved that there exists some ω -powers of infinite Borel rank. The language $h(\mathcal{V})$ is very simple to describe. It is obtained by substituting in the regular language $\mathcal{V} = (0^*.1)$ the language $\mathcal{L}.a$ to each letter $a \in \{0,1\}$, where $\mathcal{L} = \bigcup_{n \geq 1} \varphi_n(L_n)$. Notice that the language \mathcal{L} is not context free but it is the union of the increasing sequence of context free languages $\varphi_n(L_n)$. Then \mathcal{L} is a very simple recursive language and so is $h(\mathcal{V})$.

The question is left open whether there is a context free language W such that W^ω is Borel of infinite rank.

The question also naturally arises to know what are all the possible infinite Borel ranks of ω -powers of finitary languages or of finitary languages belonging to some natural class like the class of context free languages (respectively, languages accepted by stack automata, recursive languages, recursively enumerable languages, ...).

Acknowledgements. We thank the anonymous referee for useful comments on a preliminary version of this paper.

References

- [ABB96] J.-M. Autebert, J. Berstel, and L. Boasson. Context free languages and pushdown automata. In *Handbook of formal languages, Vol. 1*. Springer-Verlag, 1996.
- [Ber79] J. Berstel. *Transductions and context free languages*. Teubner Studienbücher Informatik, 1979.
- [Dup01] J. Duparc. Wadge hierarchy and veblen hierarchy: Part 1: Borel sets of finite rank. *Journal of Symbolic Logic*, 66(1):56–86, 2001.
- [Fin01] O. Finkel. Topological properties of omega context free languages. *Theoretical Computer Science*, 262(1–2):669–697, 2001.
- [Fin03] O. Finkel. Borel hierarchy and omega context free languages. *Theoretical Computer Science*, 290(3):1385–1405, 2003.
- [FS03] O. Finkel and P. Simonnet. Topology and ambiguity in omega context free languages. *Bulletin of the Belgian Mathematical Society*, 10(5):707–722, 2003.
- [HU79] J. E. Hopcroft and J. D. Ullman. *Introduction to automata theory, languages, and computation*. Addison-Wesley Publishing Co., Reading, Mass., 1979. Addison-Wesley Series in Computer Science.
- [Kec95] A. S. Kechris. *Classical descriptive set theory*. Springer-Verlag, New York, 1995.
- [Kur66] K. Kuratowski. *Topology*. Academic Press, New York, 1966.
- [Lec02] D. Lecomte. Sur les ensembles de phrases infinies constructibles a partir d’un dictionnaire sur un alphabet fini. In *Séminaire d’Initiation a l’Analyse, Vol. 1*. Université Paris 6, 2001–2002.
- [LT94] H. Lescow and W. Thomas. Logical specifications of infinite computations. In J. W. de Bakker, Willem P. de Roever, and Grzegorz Rozenberg, editors, *A Decade of Concurrency*, volume 803 of *Lecture Notes in Computer Science*, pages 583–621. Springer, 1994.
- [Mos80] Y. N. Moschovakis. *Descriptive set theory*. North-Holland Publishing Co., Amsterdam, 1980.
- [Niw90] D. Niwinski. A problem on ω -powers. In *1990 Workshop on Logics and Recognizable Sets*, University of Kiel, 1990.

- [PP04] D. Perrin and J.-E. Pin. *Infinite words, automata, semigroups, logic and games*, volume 141 of *Pure and Applied Mathematics*. Elsevier, 2004.
- [Sim92] P. Simonnet. *Automates et théorie descriptive*. PhD thesis, Université Paris VII, 1992.
- [Sta86] L. Staiger. Hierarchies of recursive ω -languages. *Elektronische Informationsverarbeitung und Kybernetik*, 22(5-6):219–241, 1986.
- [Sta97a] L. Staiger. ω -languages. In *Handbook of formal languages, Vol. 3*, pages 339–387. Springer, Berlin, 1997.
- [Sta97b] L. Staiger. On ω -power languages. In *New Trends in Formal Languages, Control, Cooperation, and Combinatorics*, volume 1218 of *Lecture Notes in Computer Science*, pages 377–393. Springer-Verlag, Berlin, 1997.
- [Tho90] W. Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, Formal models and semantics, pages 135–191. Elsevier, 1990.